SCHAEFFLER

Building reliable and trustworthy surrogate models

Dr. Dominik Penk – Schaeffler Advanced Innovation – SHARE at FAU

We pioneer motion

Personal Introduction

DR. DOMINIK PENK

Applied Research Engineer

Advanced Innovation
SHARE at FAU

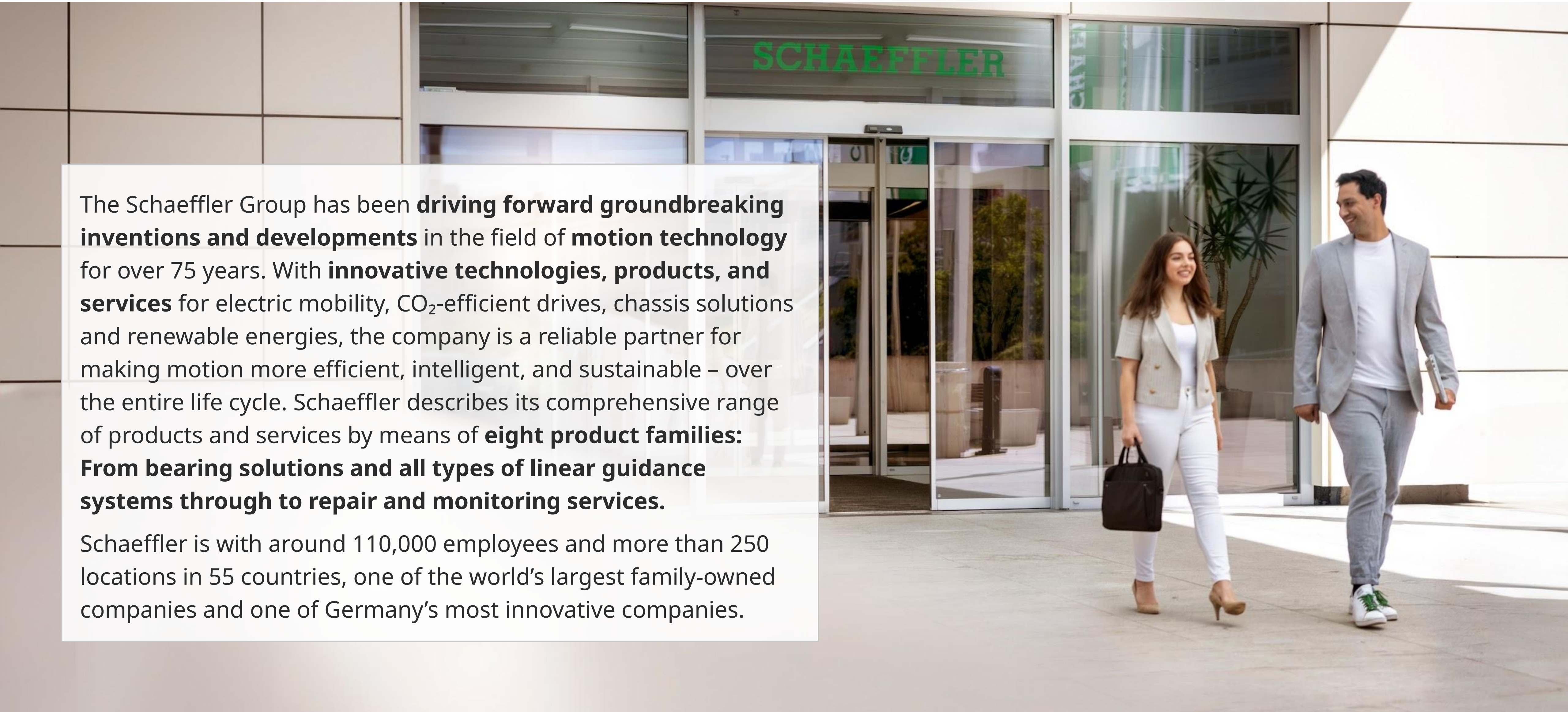
penkdmi@schaeffler.com

- **Background:**
  - 2017: master's degree in Computer Science
  - 2023: PhD in the field of computer vision – specifically for quality assessment in production
  - Joined Schaeffler and the SHARE at FAU team in August 2023

- **Role and research interest:**
  - Identifying, planning/aligning/budgeting, and managing research cooperations with focus on digitalization and ML
  - Focus on applied machine learning topics, particularly computer vision, physics-guided AI, and robust ML

SCHAEFFLER

Schaeffler Group – We pioneer motion

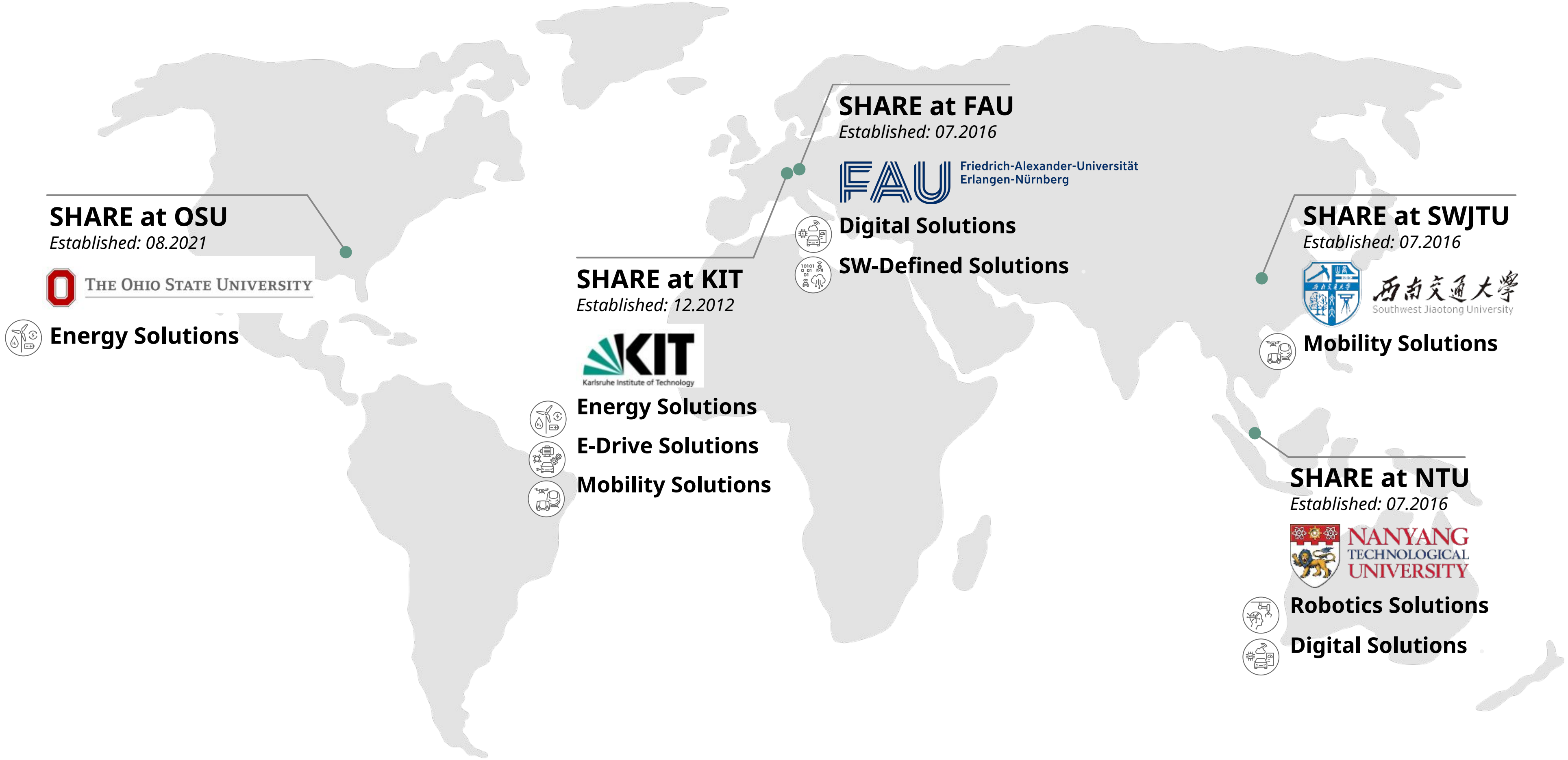


The Schaeffler Group has been **driving forward groundbreaking inventions and developments** in the field of **motion technology** for over 75 years. With **innovative technologies, products, and services** for electric mobility, CO_2-efficient drives, chassis solutions and renewable energies, the company is a reliable partner for making motion more efficient, intelligent, and sustainable – over the entire life cycle. Schaeffler describes its comprehensive range of products and services by means of **eight product families: From bearing solutions and all types of linear guidance systems through to repair and monitoring services.**

Schaeffler is with around 110,000 employees and more than 250 locations in 55 countries, one of the world's largest family-owned companies and one of Germany's most innovative companies.

SCHAEFFLER

SHARE Network



SHAREs are part of innovation clusters with own responsibility for innovative projects

SHARE at FAU Team

SCHAEFFLER

TEAM MEMBERS

Dr. Christoph Strohmeyer
06/2018

Dr. Maik Horn
04/2020

Dr. Michael Schlotter
10/2022

Dr. Dominik Riedelbauch
01/2023

Dr. Dominik Penk
08/2023

1 Use Case

2 Sobolev Training

3 Reliability Monitoring

AGEND

1 **Use Case**

2 Sobolev Training

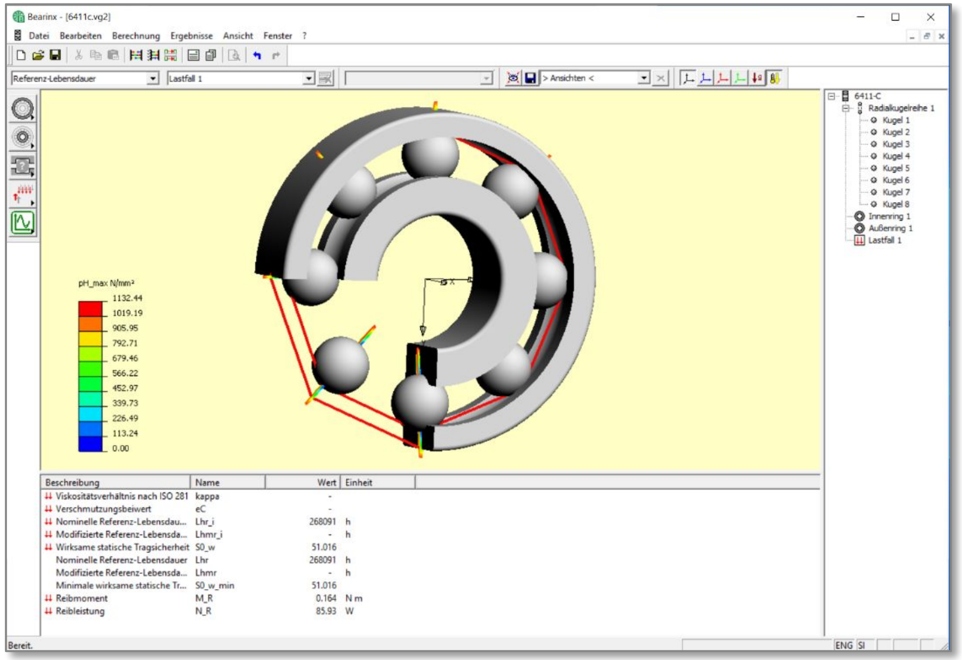
3 Reliability Monitoring

AGEND

SCHAEFFLER

Goal – Encapsulating Bearing Simulations

Bearinx – Simulation tool for roller bearings

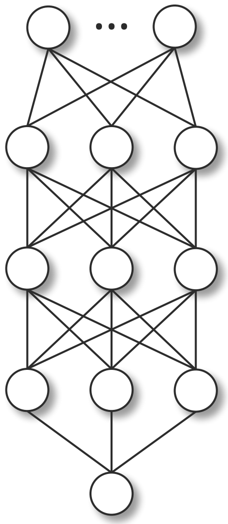


- Precise simulations based on long-standing company expertise
- Complex simulations with long simulation times (particularly for large bearings)

Data-fit Surrogate Models

Input features: load state, e.g.,

- forces, moments
- rotational speed



Prediction targets: e.g.,

- frictional torque
- equivalent load

”obfuscated” service models → **Customers**

faster component models → **System Simulation**

SCHAEFFLER

Challenges – AI in Application

Hard Challenges

Data generation costs

- Simulation slow
- Requires expert

Data sparsity

- Function to approximate; $\mathbb{R}^7 \longmapsto \mathbb{R}$
- Dataset Size: 5000

! ~ 3.4 samples per input dimension

Manifold assumption

- Does not apply
- State-of-the-art outlier detection methods do not work

Soft Requirements

Useable by non-experts

- Training handled by engineers
- Surrogate models used by engineers

Solution must be trustworthy

- Results must be correct
- Indication of uncertainty

i Training should be fully automated

i Result uncertainty must be communicated clearly

1 Use Case

2 **Sobolev Training**

3 Reliability Monitoring

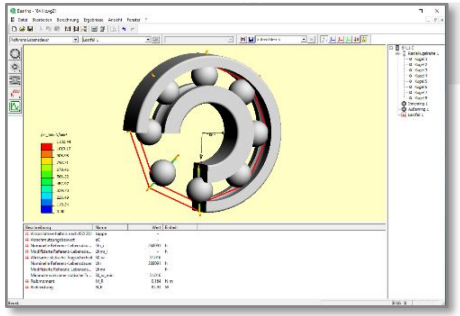
AGEND

SCHAEFFLER

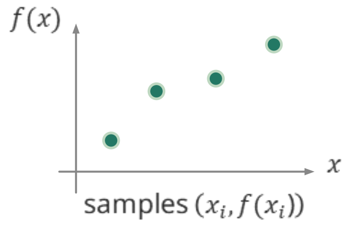
Approach

Sobolev training = Neural Network training using additional information on function gradients in each sample point [1]

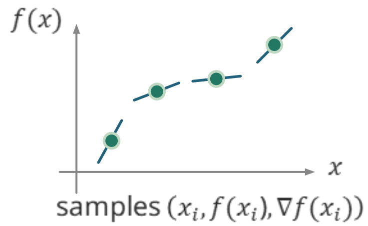
Augment data with gradient of target w.r.t. input features



Simulation result:

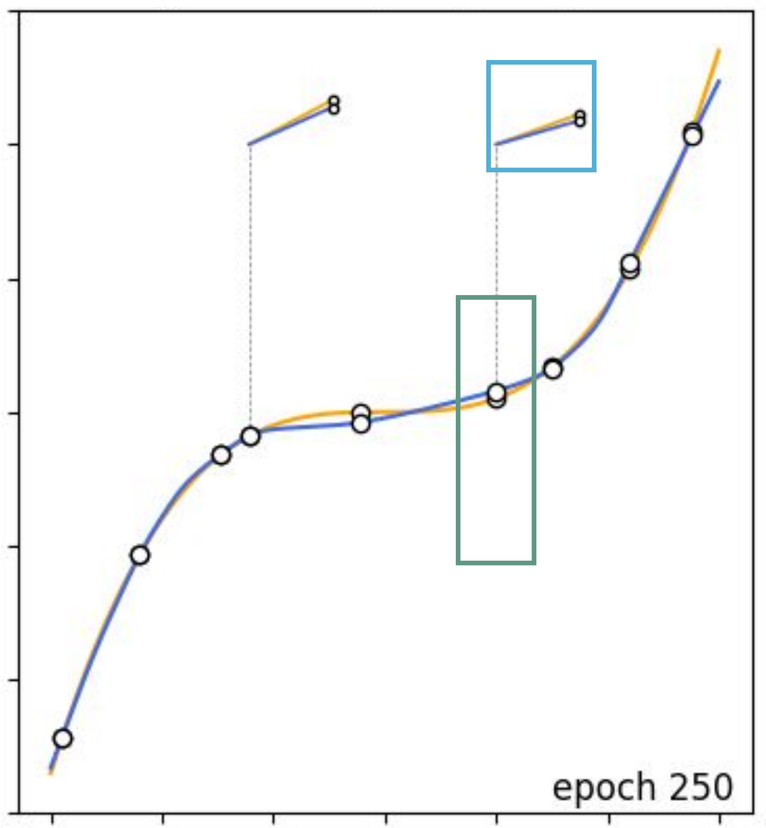


Estimate derivatives:



Train with *loss that penalizes wrong derivatives (i.e., wrong function shape!)*:

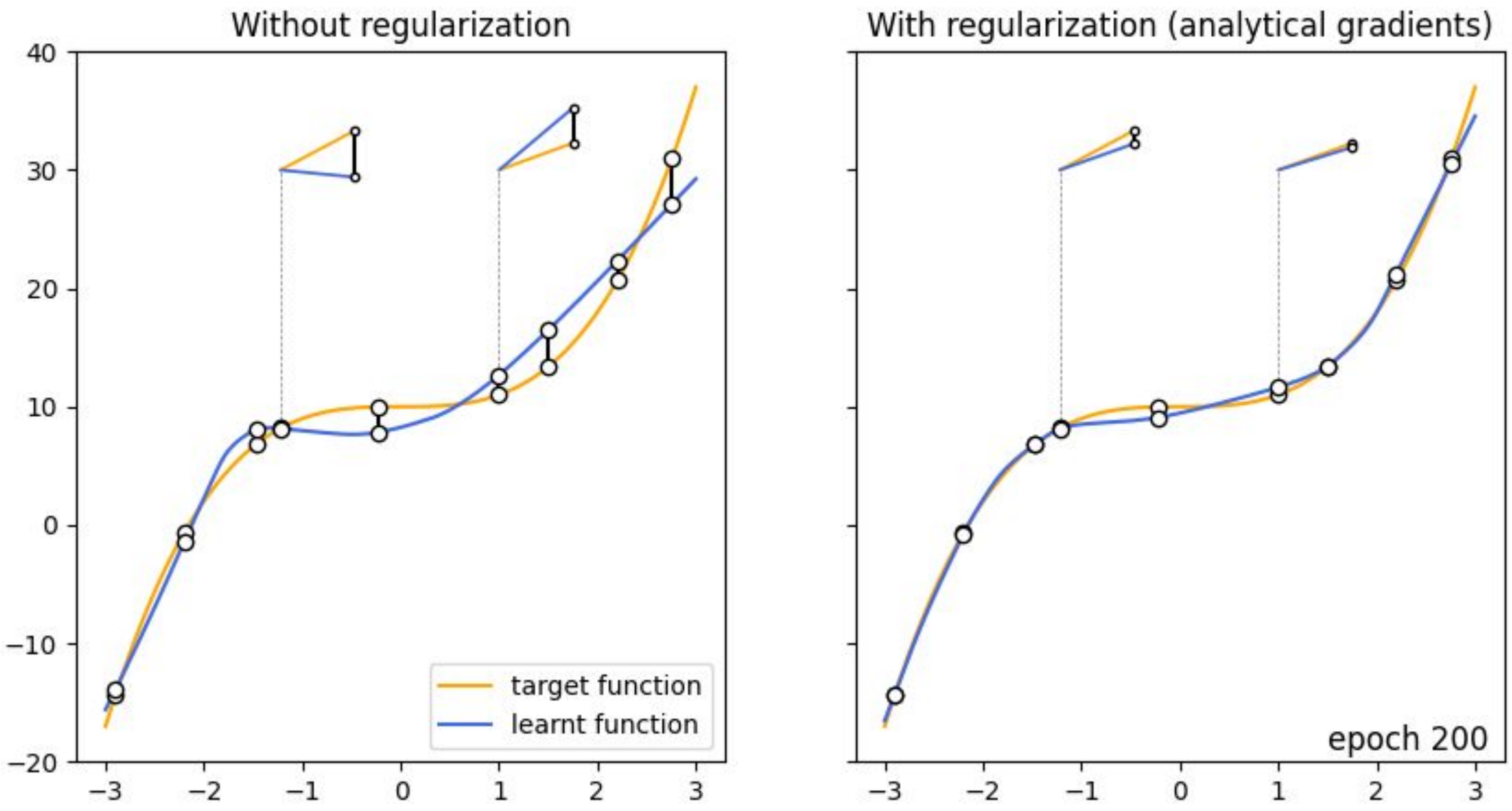
$$e_{\mathrm{ST}} = e_{\mathrm{VT}} + \beta \cdot \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\partial\, net_W(x)}{\partial x_i} - \frac{\partial\, f(x)}{\partial x_i} \right)^2$$



[1] https://link.springer.com/chapter/10.1007/978-3-030-46147-8_24

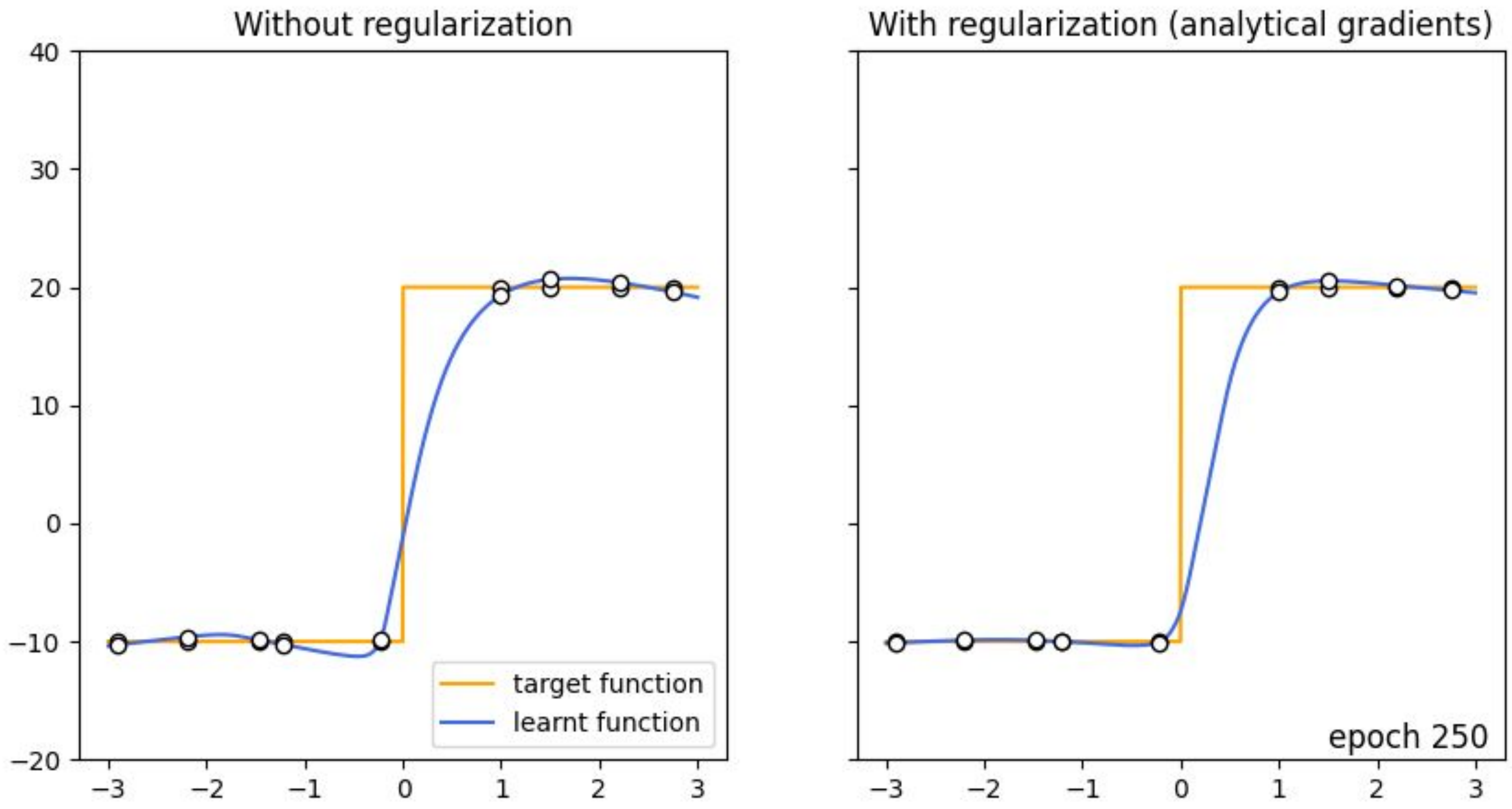
Sobolev training - Comparison

- Examples of Sobolev training effects:



Sobolev training - Comparison

– Examples of Sobolev training effects:



Insights from synthetic examples indicate plausible (and, possibly, beneficial!) effects of Sobolev training in settings with sparse data!

SCHAEFFLER

Gradient Estimation

Analytical solution

$f(x)$

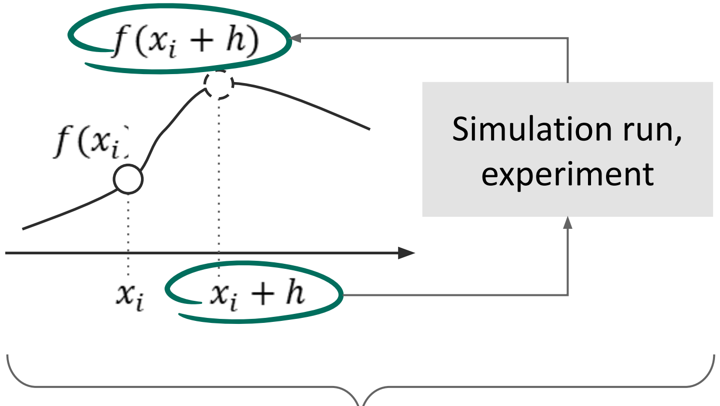
↓

Deriving the function “by hand”, analytical solvers

↓

$\nabla f(x)$

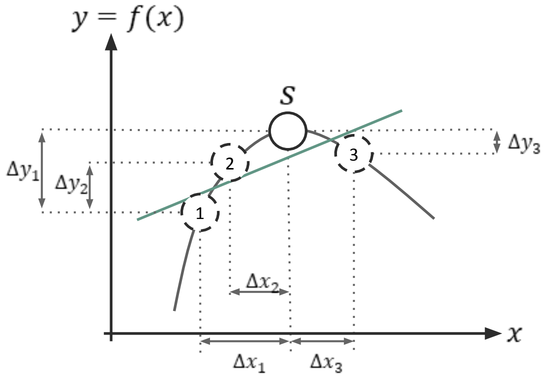
Finite differences



$$\left.\frac{\partial f}{\partial x}\right|_{x_i} \approx \frac{f(x_i + h) - f(x_i)}{h}$$

Approximation from samples

e.g., linear approximation in the neighbourhood of sample $s = (x, f(x))$



$$\left.\frac{\partial f}{\partial x}\right|_{x_i} \approx \mathrm{argmin}_d \|\Delta X \cdot d - \Delta Y\|^2$$

[1]

- + Exact information
- − Not available in most cases (and, particularly, in the case of BearinX)

- + Very precise information
- − Costly due to the required number of additional simulations runs

- + No additional samples needed
- − Approximation may be imprecise

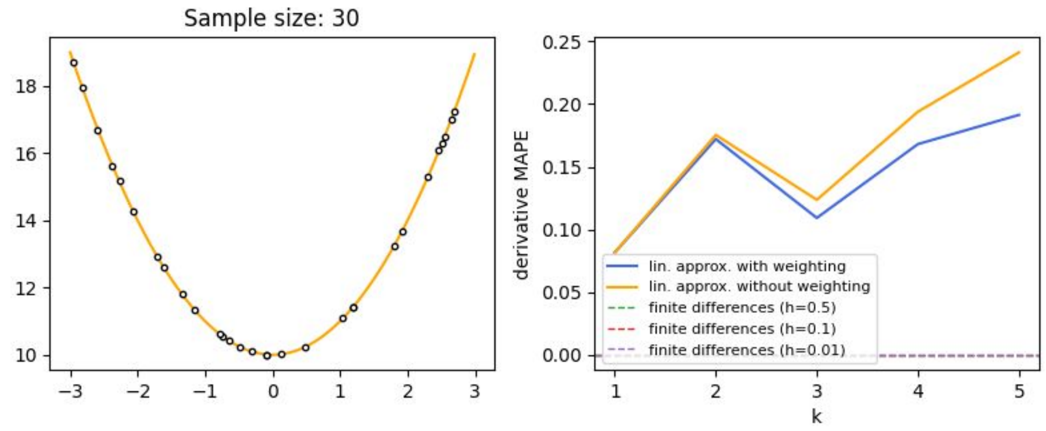
[1] https://link.springer.com/chapter/10.1007/978-3-030-46147-8_24

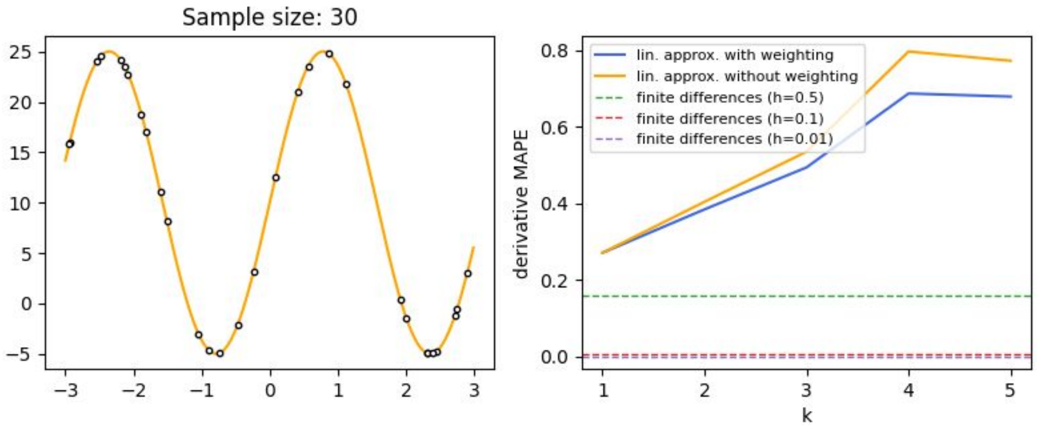
Gradient Estimation

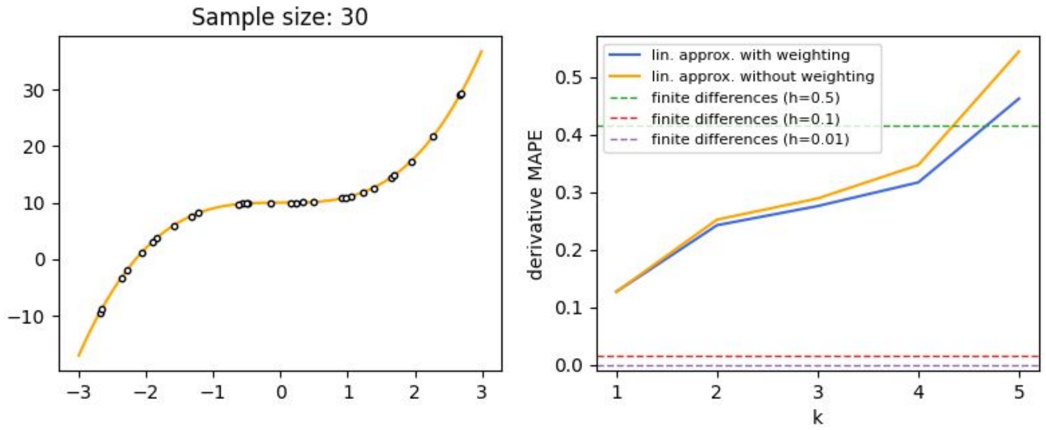
- Linear approximation of derivatives at x_i from the k nearest neighbours n_j:

$$\left.\frac{\delta f}{\delta x}\right|_{x_i} \approx \text{argmin}_d \, \|W(\Delta X \cdot d - \Delta Y)\|^2$$

  - Scaling of individual neighbours' influence by weight matrix W (influence of neighbour n_j decreases with increasing distance to x_i)
  - Analytical solution of least-squares problem

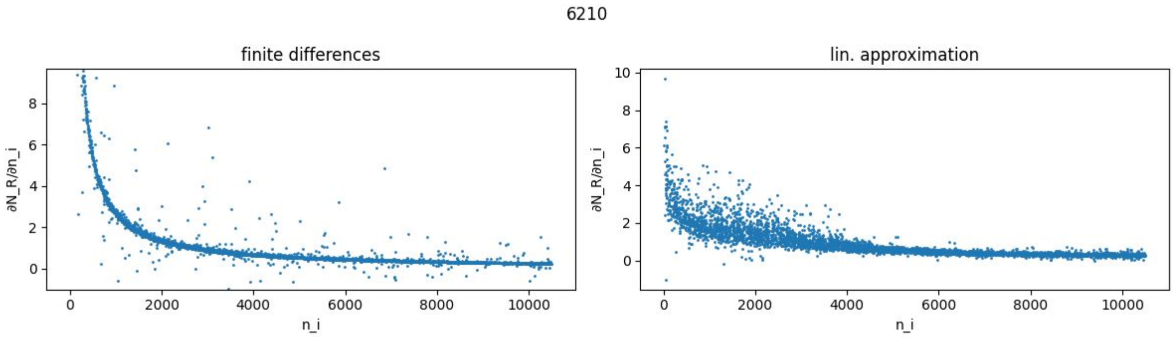


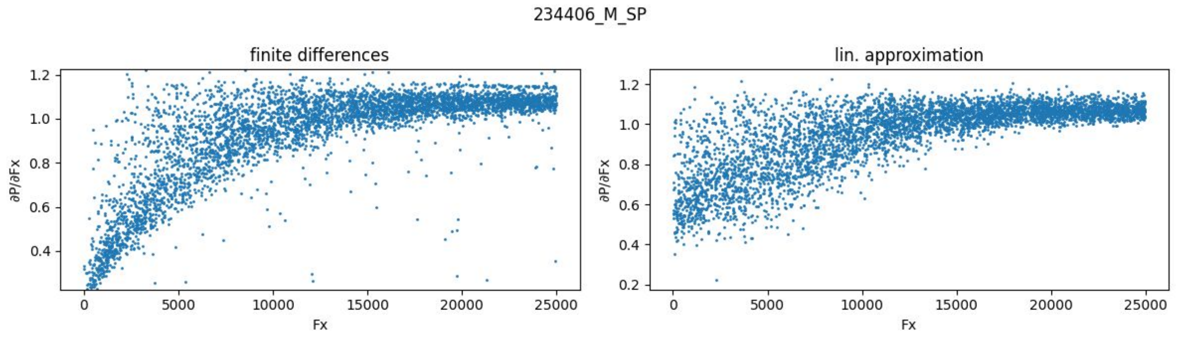


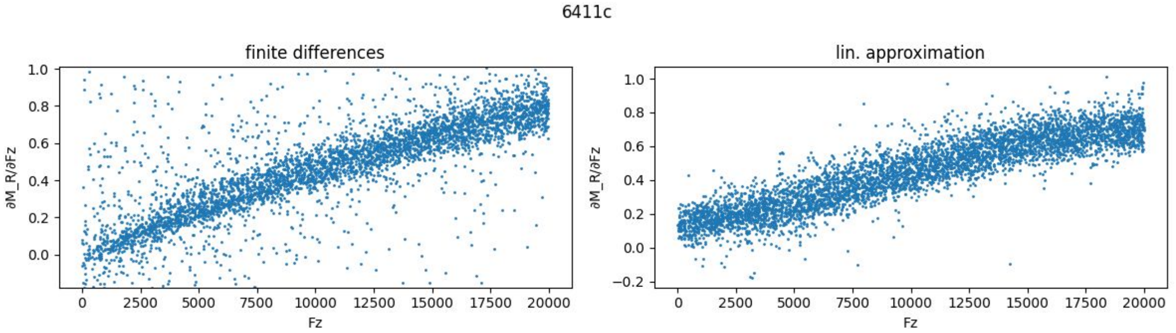


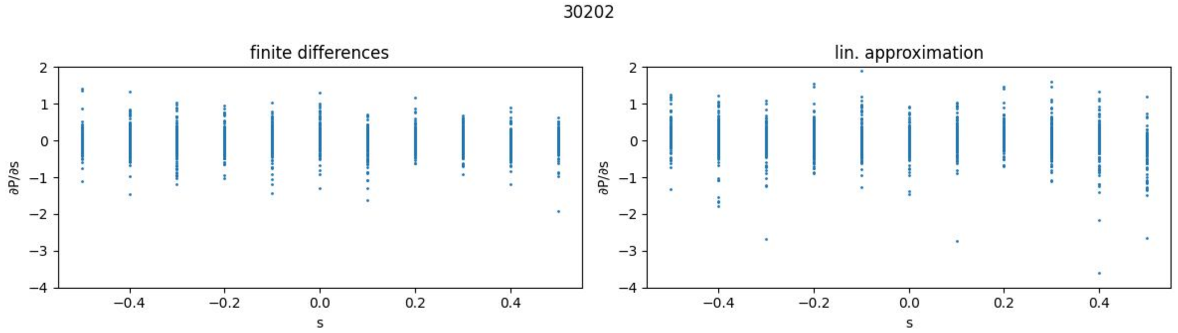
SCHAEFFLER

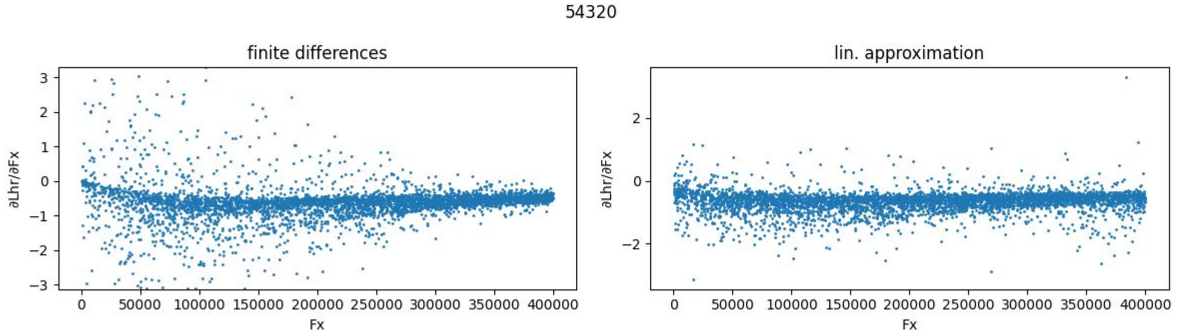
Gradient Estimation











i **Finite difference calculation for 5000 samples takes ~15 hours**

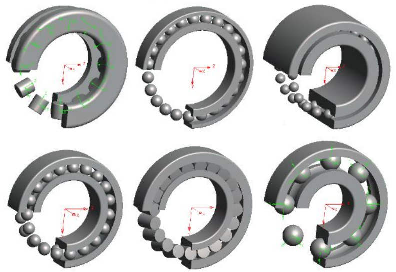
i **Linearly approximated gradients look qualitatively good/similar when compared to finite differences results**

Training Pipeline

SCHAEFFLER

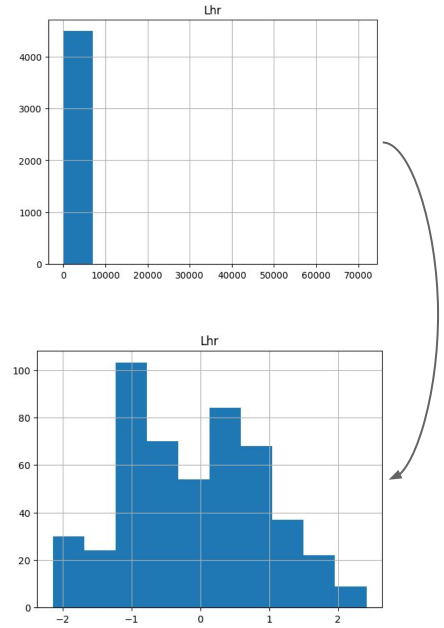
Bearing Raw Data

- Load case features:
  - F_x, F_y, F_z, M_y, M_z
  - n_i (rpm)
  - s
- Prediction targets:
  - $P0$
  - N_R
  - M_R
  - Lhr



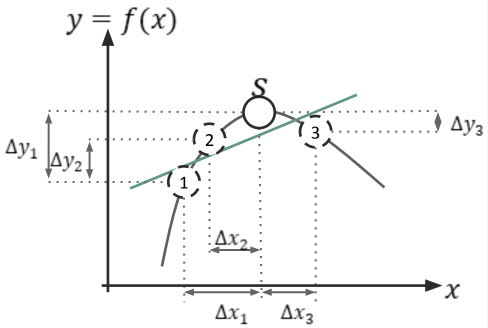
Data Preprocessing

- un-skewing
- standardization



Gradient Estimation

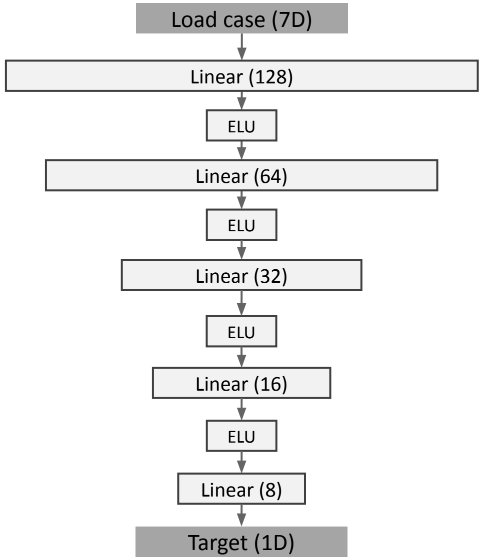
- Linear approximation from the available samples



$$\left.\frac{\delta f}{\delta x}\right|_{x_i} \approx \text{argmin}_d \|W(\Delta X \cdot d - \Delta Y)\|^2$$

Model Training

- Relatively small neural network architecture
- **Derivative loss component fades out over time**



SCHAEFFLER

Improvements on small datasets

Training of neural networks for all bearings and targets with Sobolev regularization based on linear gradient estimates (5.000 samples)

- Training hyperparameters in line with baseline training
- Parameter study over regularization strength parameter β

$$e_{ST} = e_{VT} + \boxed{\beta} \cdot \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\partial\, net_W(x)}{\partial x_i} - \frac{\partial\, f(x)}{\partial x_i} \right)^2$$

Sobolev training can decrease errors compared to baseline

In absolute numbers, gains due to regularization are small

Regularization strength is yet another hyperparameter to tune

Improvement by regularization

	Lhr	M_R	N_R	P
234406_M_SP	36.7%	5.8%	4.1%	0.9%
30202_A	8.4%	-1.7%	2.4%	-7.2%
54320	-6.5%	0.8%	13.5%	6.3%
6210	15.0%	1.7%	7.0%	6.3%
6411c	-1.1%	12.9%	21.6%	1.1%

Best regularization strength

	Lhr	M_R	N_R	P
234406_M_SP	$\beta = 1000.0$	$\beta = 10.0$	$\beta = 1.0$	$\beta = 1.0$
30202_A	$\beta = 1.0$	$\beta = 1000.0$	$\beta = 0.1$	$\beta = 100.0$
54320	$\beta = 1.0$	$\beta = 100.0$	$\beta = 10.0$	$\beta = 1.0$
6210	$\beta = 10.0$	$\beta = 10.0$	$\beta = 10.0$	$\beta = 1000.0$
6411c	$\beta = 100.0$	$\beta = 100.0$	$\beta = 1000.0$	$\beta = 100.0$

SCHAEFFLER

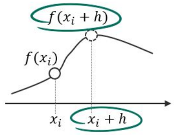
Influence of gradient regularization vs. data set size

Comparing the effects of Sobolev training with finite differences gradient estimates vs. linear gradient estimates

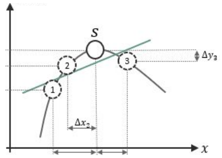
- Finite (forward) differences gradient estimate for one sample with 7 features requires 7 additional simulation runs (one perturbation per feature), i.e.:

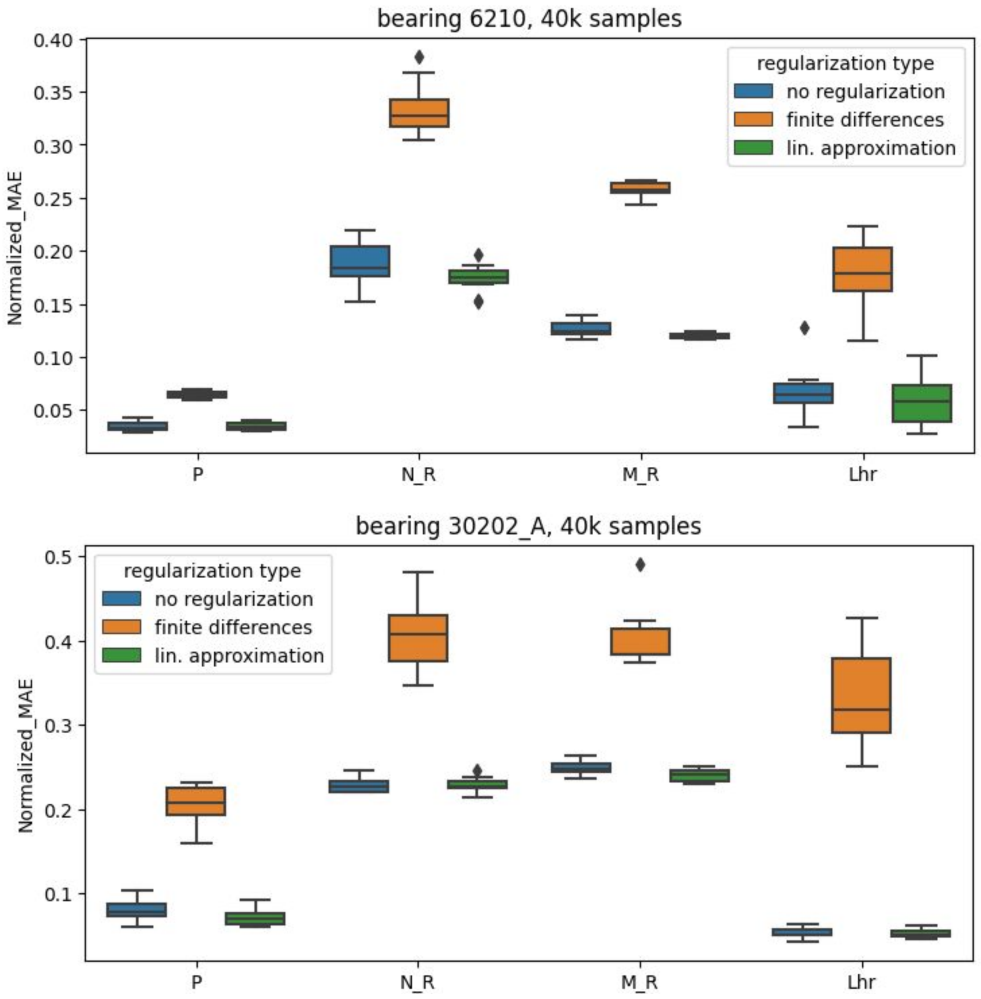
$$5000 \text{ samples with FD gradients} \longleftrightarrow 40000 \text{ simulation runs}$$

- For a fair comparison, we trained on



5k samples with FD gradients **vs.** *40k samples (uniform random) w(/o) linear grad. estimates*



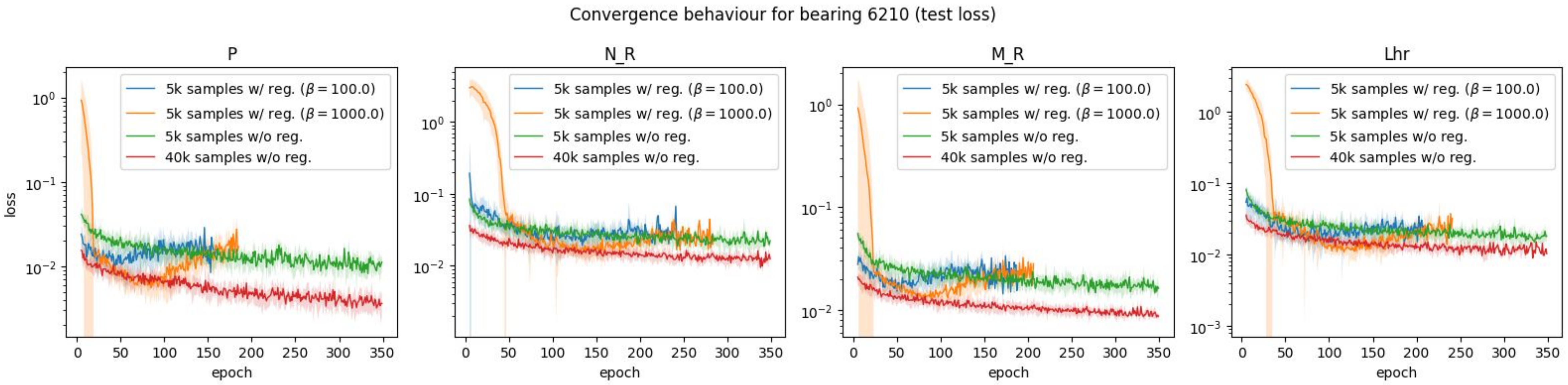


i **We could not observe gains due to regularization with linear gradient estimates at larger dataset sizes**

i **Instead of condensing a large dataset into a finite differences gradient estimate, one can as well train without regularization on a uniform random sample of the same size**

SCHAEFFLER

Influence on training convergence



In theory, Sobolev regularization with precise finite differences gradient estimates can speed up convergence

In our case, this speed-up is over-compensated by the duration of additional simulation queries (cf. considerations on fair comparison)

effect observed for at least one target:

	FD gradients	lin. approx. gradients
6210	✓	✗
6411c	✓	✗
30202-A	✗	✗
54320	✓	✗
234406-M-SP	✓	✗

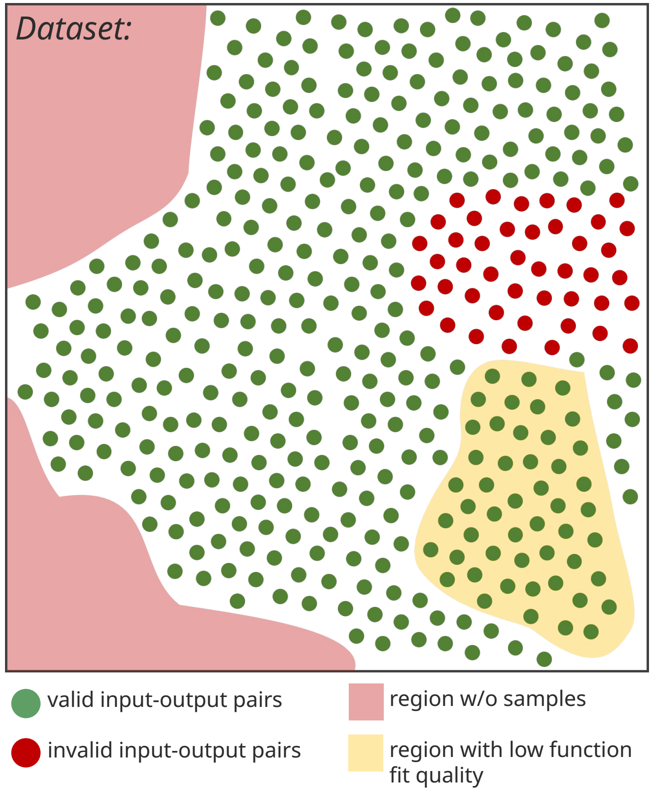
1 Use Case

2 Sobolev Training

3 Reliability Monitoring

AGEND
A

Safeguarding with a "Neural Network Traffic Light"



1

Data Safeguarding = Monitoring for model inputs outside the training data domain

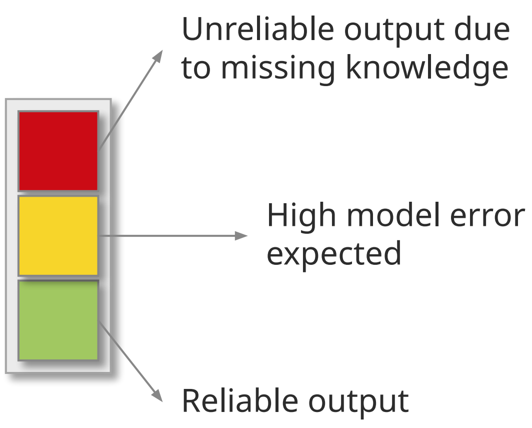
Queries in input areas …

- without available training data
- that were only sampled sparsely
- where the simulation did not converge

2

Model Safeguarding = Monitoring for model inputs in regions where the model does not fit well

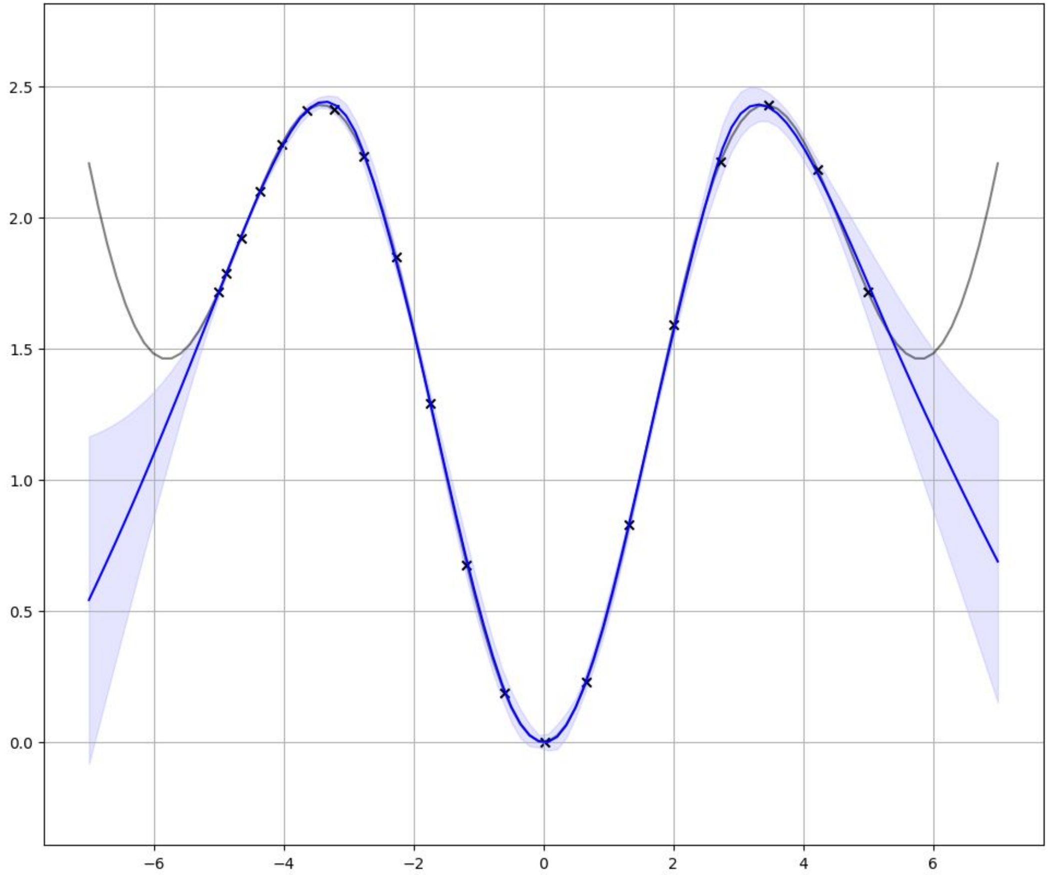
- Queries in input areas where the model is expected to over- or underfit



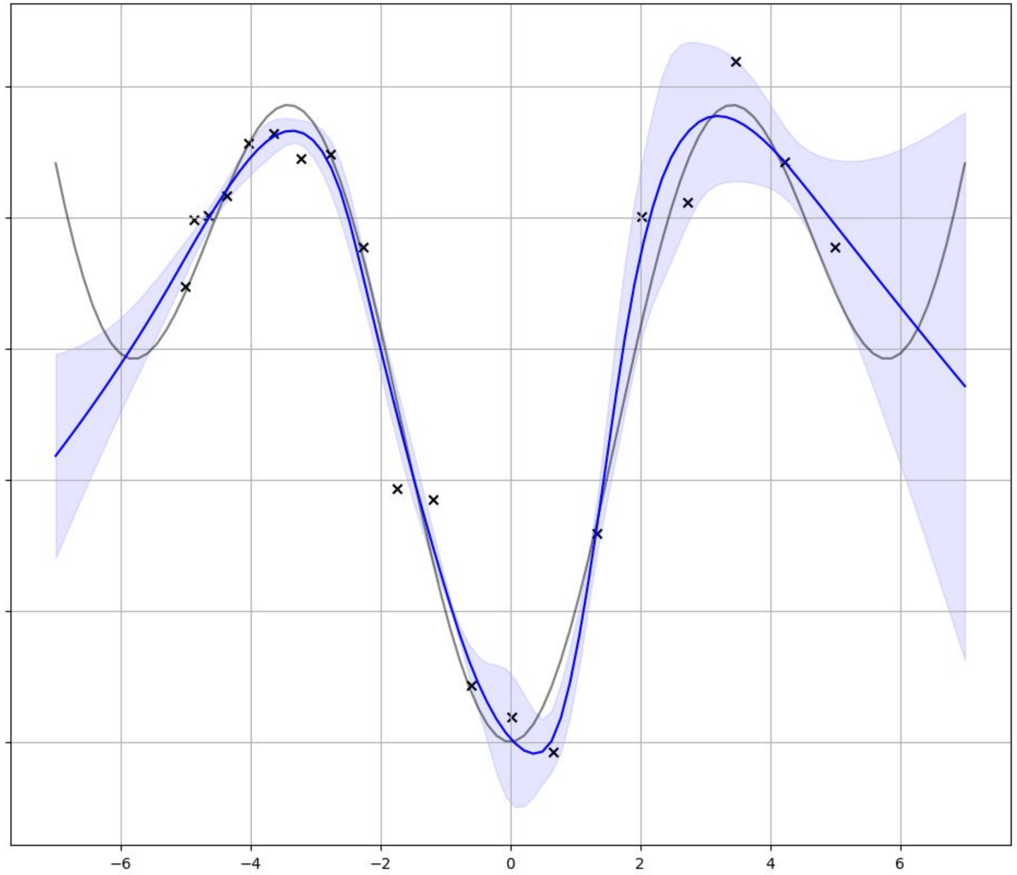
Our approach is built to safeguard models independently of the underlying data sampling strategy!

Approach - Ensemble Networks

Noise Free Data



Noisy Data



Ensemble do not work for noise-free data

SCHAEFFLER

Data Safeguarding

1 **Data Safeguarding** = Monitoring for model inputs outside the training data domain

Queried point closer to non-converged than to converged samples? → yes

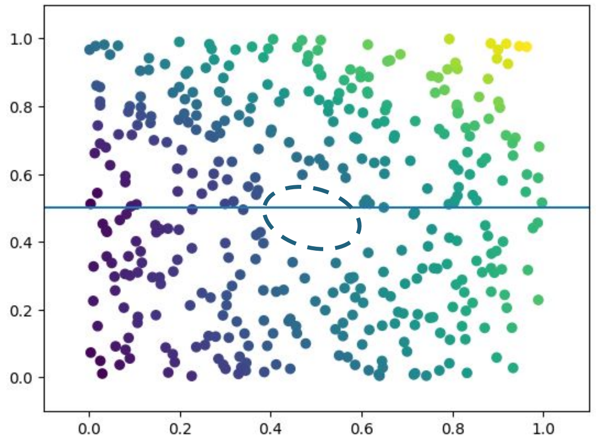
Distance between queried point and next neighbours (local density) smaller than threshold? → yes

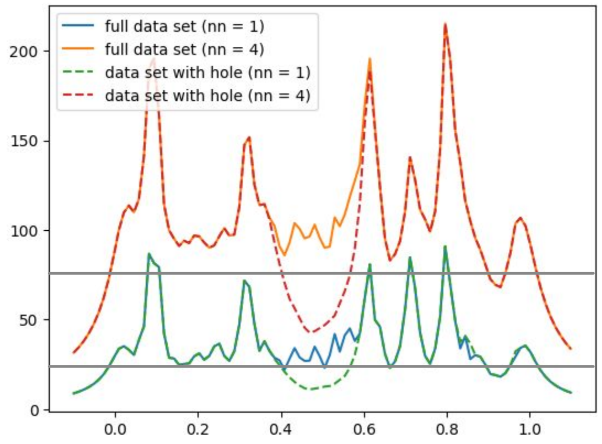
Model safeguarding

Calculating **meaningful distances (and densities)** [1]:

- A queried point may be far away from the training domain regarding one of the features
- This is not necessarily a problem as long as that feature does not correlate with the regression target!
- Therefore: **vicinity score** = correlation-weighted Euclidean distance

$$d_{ij} = \sqrt{\sum_{k=1}^{n} w_{k'} \left(x'_{k,i} - x'_{k,j}\right)^2} \qquad S_i = \sum_{j=1}^{N} 1/d_{ij}.$$

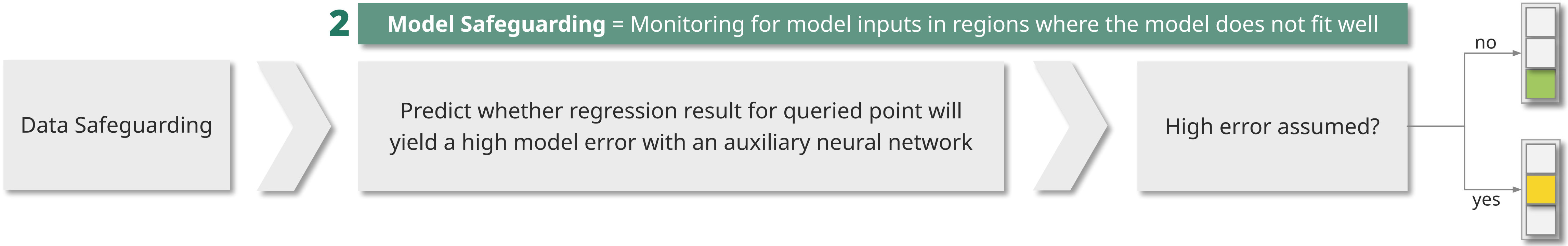




[1] Evan Askanazi and Ilya Grinberg, 2024 Mach. Learn.: Sci. Technol. 5 025030, “Analysis of machine learning prediction reliability based on sampling distance evaluation with feature decorrelation”

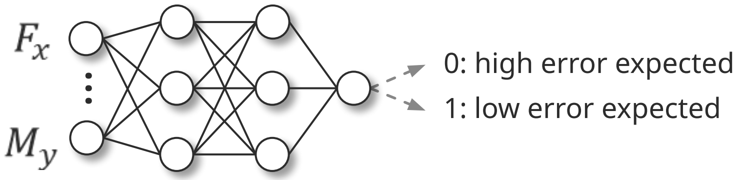
SCHAEFFLER

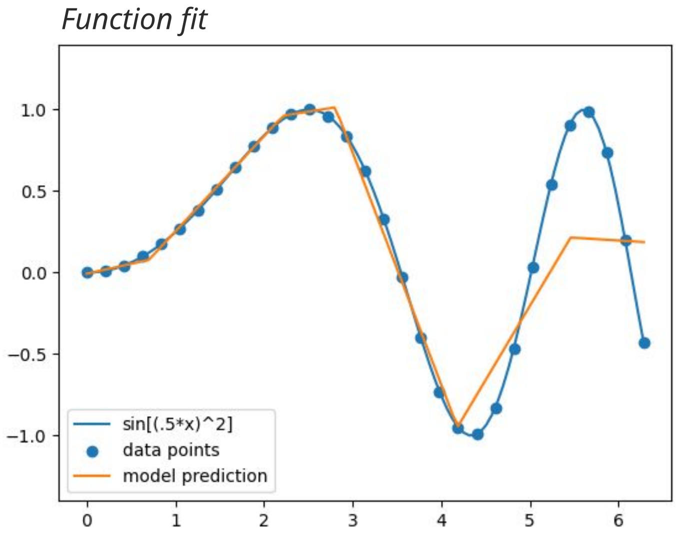
Model Safeguarding

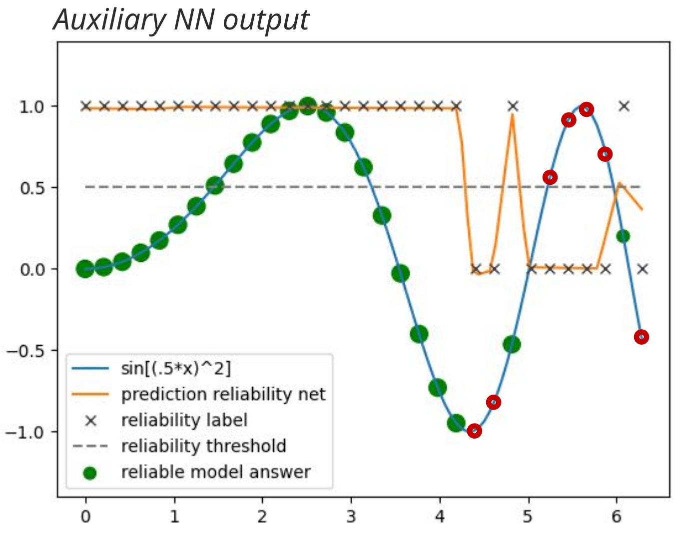


Training an **auxiliary neural network** for error prediction:

- Calculate sample-wise absolute error of regression target for the original dataset
- Binarize sample-wise errors with **a user-defined error threshold** □ new dataset with load case to 0/1 label pairs
- Train a classifier on the “new” dataset:







SCHAEFFLER

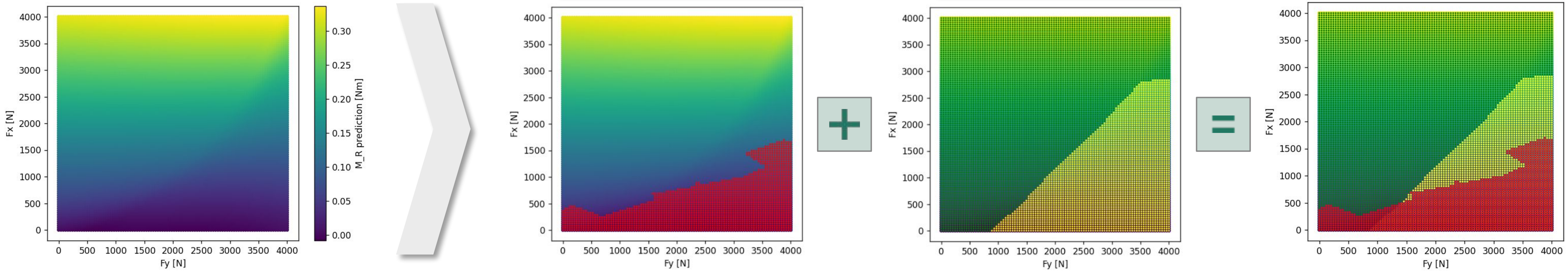
Results

Qualitative example of 2-step reliability monitoring

Frictional torque prediction* w/o reliability monitoring

1 Data safeguarding

2 Model safeguarding



Quantitative evaluation of model safeguarding for all bearings and targets (dataset size: 5.000 samples, 10-fold cross-validation)

- Error threshold: 25% quantile of error distribution on training set
- Unreliable points were excluded from regression MAE calculation (as if falling back to the Bearinx simulation in these cases)

Safeguarding can detect unreliable regions of surrogate models successfully

Improvement of MAE by reliability prediction

bearing \ target	Lhr	M_R	N_R	P	P0
1219-M_5k	61.83%	59.36%	56.32%		72.39%
2217-M_5k	98.86%	78.72%	68.17%		63.43%
234406-M-SP_5k	92.97%	11.37%	17.59%	30.38%	14.04%
30202-A_40k	97.37%	42.80%	29.80%	76.72%	54.93%
30202-A_5k	98.38%	33.69%	12.60%	78.48%	42.37%
54320_5k	93.32%	22.64%	24.25%	32.37%	
6210_40k	93.64%	7.88%	50.65%	64.58%	30.08%
6210_5k	97.04%	28.57%	30.50%	44.19%	29.11%
6411-C_5k	98.59%	48.28%	39.53%	51.63%	20.13%

* Bearing 30202-A, operating point $n_i = 3000\ rpm,\ F_z = 0\ N,\ M_y = M_z = 0\ Nm,\ s = 0\ \mu m$

SCHAEFFLER